Abstract

The current state of technology can positively affect how the internet is used as well as how information, particularly information on the film industry, is disseminated online. Many movie reviews can be simply found thanks to this ease. Reviews of movies have a big impact in the various ways movies are available. Thanks to the ease of various information on the internet, the number of movie reviews has become diverse. Consequently, conducting a sentiment analysis is required. In this research, classification method used is Logistic Regression. The method was chosen because it has accurate classification accuracy. In this study, Information Gain was also chosen as a feature selection because it is good enough to do a filter approach in classification. Furthermore, for feature extraction, TF-IDF was chosen because it can overcome data imbalance in the dataset. The best model resulting from this research is a model built without using stemming in the preprocessing stage, without using information gain feature selection, and using parameters in Logistic Regression which produces an f1-score of 76.50%.

Keywords: Movie Review, Sentiment Analysis, Information Gain, Logistic Regression